



Attorney's Docket No. 16-358

Patent

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the

Application of: SURAJIT CHAUDHURI, ET AL. Examiner: N/A

Application No.: 10/600,083 Art Unit:

Filed: JUNE 20, 2003

Title: EFFICIENT FUZZY MATCH FOR  
EVALUATING DATA RECORDS

### INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Enclosed is a copy of Information Disclosure Citation Form PTO-1449 together with a copy of the document(s) listed on that form. Applicant respectfully requests that the Examiner consider the listed document(s), initial the enclosed Form PTO-1449 to indicate such consideration, and return a copy of the initialed Form PTO-1449 to the Applicant.

This Information Disclosure Statement is being submitted pursuant to 37 C.F.R. § 1.97(b).

I hereby certify that this paper is being deposited today with the U.S. Postal Service as 1st Class Mail addressed to the Assistant Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

on August 13, 2003

By: [Signature]

-1-

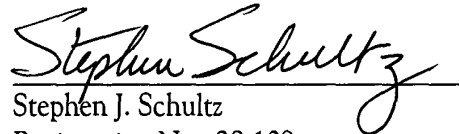
Serial No. 10/600,083  
Attorney Docket No. 16-358

Pursuant to 37 C.F.R. § 1.97, the submission of this Information Disclosure Statement shall not be construed to be an admission that the information cited in this Information Disclosure Statement is, or is considered to be, material to patentability.

The Commissioner is hereby authorized to charge any required fee under 37 C.F.R. § 1.17 in connection with this communication to our Deposit Account No. 23-0630.

Respectfully submitted,

Date: August 12, 2003

  
Stephen J. Schultz  
Registration No. 29,108

WATTS, HOFFMANN, FISHER  
& HEINKE CO., L.P.A.  
1100 Superior Ave., Ste. 1750  
Cleveland, Ohio 44114

Phone: (216) 241-6700  
Facsimile: (216) 241-8151



Form PTO-1449 U.S. DEPARTMENT OF COMMERCE (Modified) Patent and Trademark Office				Attorney's Docket Number 16-358		Application Number 10/600,083	
INFORMATION DISCLOSURE CITATION				Applicant Surajit Chaudhuri, et al.		Filed June 20, 2003	
				Examiner		Art Unit	
Sheet 1 of 2							
U.S. PATENT DOCUMENTS							
EXAMINER INITIALS	DOCUMENT NUMBER	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE	
FOREIGN PATENT DOCUMENTS							
	DOCUMENT NUMBER	DATE	COUNTRY	CLASS	SUBCLASS	TRANSLATION	
						YES	NO
OTHER DOCUMENTS							
	R. Ananthakrishna, S. Chaudhuri and V. Ganti. <u>Eliminating Fuzzy Duplicates in Data Warehouses</u> . In <i>Proceedings of VLDB</i> , Hong Kong, 2002.						
	A.N. Arslan, O. Egecioglu and P.A. Pevzner. <u>A New Approach to Sequence Comparison: Normalized Local Alignment</u> . <i>Bioinformatics</i> . 17(4):327-337, 2001						
	J.A. Aslam, K. Pelekhev and D. Rus. <u>Static and Dynamic Information Organization with Star Clusters</u> . <i>CIKM</i> 1998, pages 208-217.						
	J.A. Aslam, K. Pelekhev and D. Rus. <u>A Practical Clustering Algorithm for Static and Dynamic Information Organization</u> . <i>ACM-SIAM Symposium on Discrete Algorithms</i> , 1999.						
	V. Borkar, K. Deshmukh and S. Sarawagi. <u>Automatic Segmentation of Text Into Structured Records</u> . In <i>Proceedings of ACM SIGMOD Conference</i> , Santa Barbara, CA, May 2001.						
	A. Broder, S. Glassman, M. Manasse and G. Zweig. <u>Syntactic Clustering of the Web</u> . In <i>Proc. Sixth Int'l. World Wide Web Conference, World Wide Web Consortium</i> , Cambridge, pages 391-404, 1997.						
	K. Beyer, J. Goldstein, R. Ramakrishnan and U. Shaft. <u>When is "Nearest Neighbor" Meaningful?</u> <i>International Conference on Database Theory</i> , pages 217-235. January 1999.						
	S. Chaudhuri, K. Ganjam, V. Ganti and R. Motwani. <u>Robust and Efficient Fuzzy Match for Online Data Cleaning</u> . In <i>Proceedings of ACM SIGMOD</i> , San Diego, CA June 2003.						
	W. Cohen. <u>Integration of Heterogenous Databases Without Common Domains Using Queries Based on Textual Similarity</u> . In <i>Proceedings of ACM SIGMOD</i> , pages 201-212, Seattle, WA June 1998.						
	R. Forino. <u>Data e-quality: A behind the Scenes Perspective on Data Cleansing</u> . <a href="http://www.dmreview.com/">http://www.dmreview.com/</a> , March 2001.						
	H. Galhardas, D. Florescu, D. Shasha, E. Simon and C. Saita. <u>Declarative Data Cleaning: Language, Model, and Algorithms</u> . In <i>Proceedings of the 27th International Conference on Very Large Databases</i> , pages 371-380, Roma, Italy, September 11-14, 2001.						
	H. Galhardas, D. Florescu, D. Shasha and E. Simon. <u>An Extensible Framework for Data Cleaning</u> . In <i>ACM SIGMOD</i> , May 1999.						
	L. Gravano, P. Ipeirotis, H.V. Jagadish, N. Kousas, S. Muthukrishnan and D. Srivastava. <u>Approximate String Joins in a Database (Almost) For Free</u> . In <i>Proceedings of the VLDB</i> 2001.						
	V. Ganti, J. Gehrke and R. Ramakrishnan. <u>Cactus-Clustering Categorical Data Using Summaries</u> . In <i>Proceedings of the ACM SIGKDD Fifth International Conference on Knowledge Discovery in Databases</i> , pages 73-83, August 15-18, 1999						
	D. Gibson, J. Kleinberg and P. Raghavan. <u>Clustering Categorical Data: An Approach Based on Dynamical Systems</u> , <i>VLDB</i> 1998, New York City, New York, August 24-27						
	S. Guha, R. Rastogi and K. Shim. <u>Rock: A Robust Clustering Algorithm for Categorical Attributes</u> . In <i>Proceedings of the IEEE International Conference on Data Engineering</i> , Sydney, March 1999.						
	Y. Huhtala, J. Karkkainen, P. Porkka and H. Toivonen. <u>Efficient Discovery of Functional and Approximate Dependencies Using Partitions</u> . In <i>Proceedings of the 14th International Conference on Data Engineering (ICDE)</i> , pages 392-401, Orlando, Florida, February 1998.						
EXAMINER				DATE CONSIDERED			
EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP § 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.							

